Big* Computing

UIC Alumni Colloquium

Burt Holzman (class of 2000)

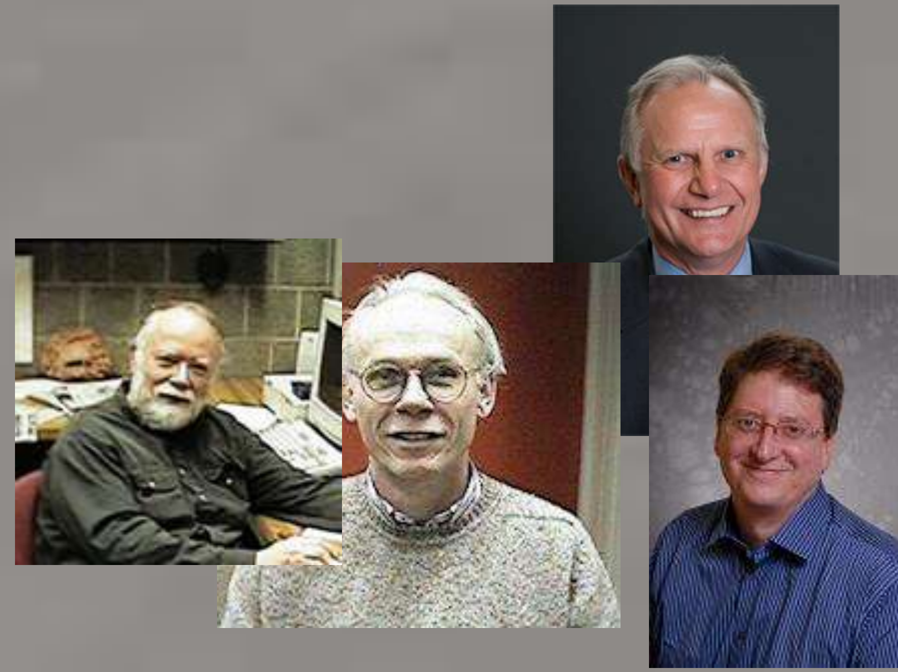
* For some arbitrary value of "Big"

Ten years in one minute

- University of Illinois at Chicago
  - Graduate Student / Slave 1995-2000
  - PhD, Nuclear Physics, **E917 experiment**
- Brookhaven National Laboratory
  - Post-doctoral Research Associate 2001-2004
  - Assistant Scientist 2004-2005
  - Head of Computing, **PHOBOS Experiment**
- Fermi National Accelerator Laboratory
  - Computing Professional 2005-present
  - US CMS Grid Services & Interfaces Coordinator
  - CMS Computing Facilities Head @ FNAL , **CMS Experiment**

UIC Colleagues

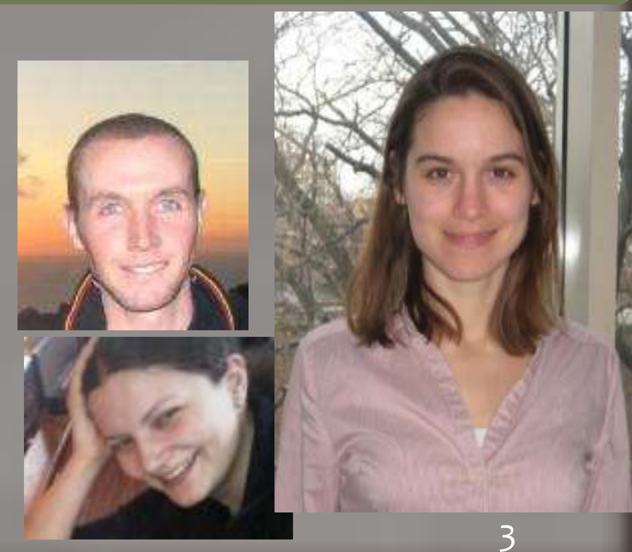
Postdocs



Faculty

Grad Students

Undergrads



E917 Collaboration

Argonne National Laboratory

B. B. Back, R. R. Betts, A. Gillitzer, **D. Hofman**, V. Nanal, A. H. Wuosmaa

Brookhaven National Laboratory

Y. Y. Chu, J. Cumming

University of California at Riverside

J. Chang, W. C. Chang, W. Eldredge, S. Y. Fung, **R. Seto**, H. Wang, H. Xiang, G. Xu, C. Zou

Columbia University

C. Y. Chi, M. Moulson

University of Illinois at Chicago

R. R. Betts, **B. Holzman**, **R. Ganz**, **D. McLeod**

University of Maryland

E. Garcia, A. C. Mignerey, D. Russ, P. J. Stanskas, A. Ruangma

Massachusetts Institute of Technology

J. C. Dunlop, G. Heintzelman, C. A. Ogilvie, G. S. F. Stephans, H. B. Yao

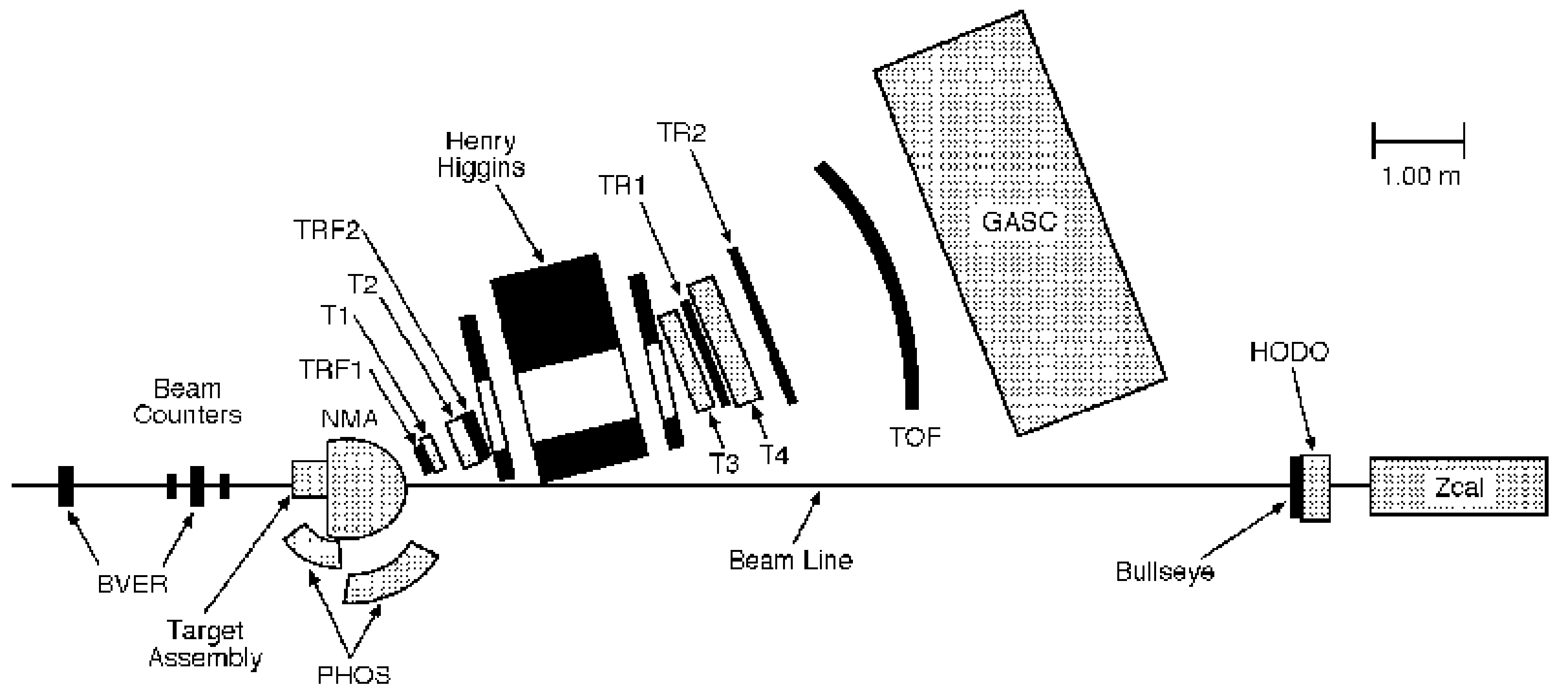
University of Rochester

R. Pak, F. L. H. Wolfs

Yonsei University, Korea

J. H. Kang, E. J. Kim, S. Y. Kim, Y. Kwon

E917 – my thesis experiment



- Experiment E917 @ AGS (fixed target)
- Run dates: 11/5/96 – 1/31/97
- Target, Projectile: Au+Au
- E_{BEAM} = 6, 8, 10.8 GeV/u
- Movable spectrometer: $.57 < \eta < 2.10$

E917 Physics

- **Phi meson mass and yield**
  - Motivated by earlier indications that phi mass may have shifted – one of the predicted signals of a quark-gluon plasma
- Measured excitation functions of kaons, pions
- Proton yields and spectra
- Particle ratios in thermal models
- Directed (v1) and elliptic (v2) proton and pion flow
- Two-particle correlations

Total number of citable papers analyzed:	15
Total number of citations:	364
Average citations per paper:	24.3
Renowned papers (500+)	0
Famous papers (250-499)	0
Very well-known papers (100-249)	0
Well-known papers (50-99)	3
Known papers (10-49)	4
Less known papers (1-9)	7
Unknown papers (0)	1
h-index	7

Citation data from INSPIRE

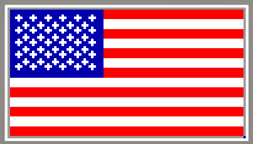
E917 Computing

- 300 million events, 15 kB each
- Aggregate data size: 5 TB
- Full dataset reconstruction pass: 3 years

- Data reconstructed on a few DEC Alphas located at Maryland, BNL, UC-Riverside, Rochester, Argonne (early example of distributed computing?)
- (upgraded to 128 MB of memory and 23 GB disks in 97-98!)

PHOBOS Collaboration (July 2006)

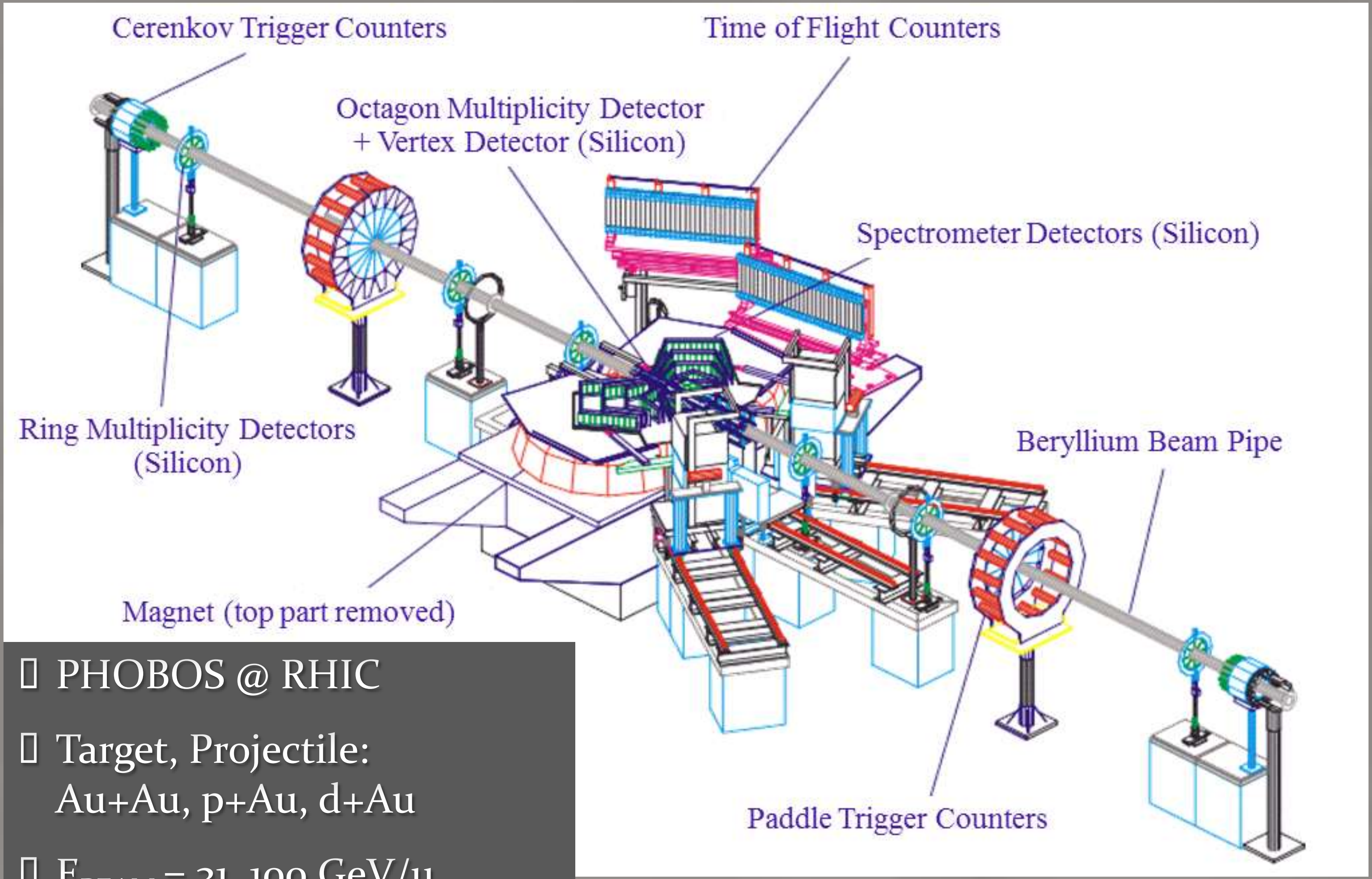




Burak Alver, Birger Back, Mark Baker, Maarten Ballintijn, Donald Barton, **Russell Betts**, Richard Bindel, Wit Busza (Spokesperson), **Vasundhara Chetluru**, **Edmundo García**, Tomasz Gburek, Joshua Hamblen, Conor Henderson, **David Hofman**, **Richard Hollis**, Roman Hołyński, **Burt Holzman**, **Aneta Iordanova**, Chia Ming Kuo, Wei Li, Willis Lin, Constantin Loizides, Steven Manly, Alice Mignerey, Gerrit van Nieuwenhuizen, **Rachid Nouicer**, Andrzej Olszewski, Robert Pak, Corey Reed, Christof Roland, Gunther Roland, **Joe Sagerer**, Peter Steinberg, George Stephans, Andrei Sukhanov, Marguerite Belt Tonjes, Adam Trzupek, Sergei Vaurynovich, Robin Verdier, Gábor Veres, Peter Walters, Edward Wenger, Frank Wolfs, Barbara Wosiek, Krzysztof Woźniak, Bolek Wysłouch

ARGONNE NATIONAL LABORATORY
INSTITUTE OF NUCLEAR PHYSICS PAN, KRAKOW
NATIONAL CENTRAL UNIVERSITY, TAIWAN
UNIVERSITY OF MARYLAND
BROOKHAVEN NATIONAL LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
UNIVERSITY OF ILLINOIS AT CHICAGO
UNIVERSITY OF ROCHESTER

PHOBOS Experiment



- PHOBOS @ RHIC
- Target, Projectile: Au+Au, p+Au, d+Au
- E_{BEAM} = 31, 100 GeV/u

Phobos Physics

- **Charged-particle multiplicity**
  - Predictions for RHIC differed by more than an order of magnitude!
- Antiparticle/particle ratios
- Two (and more) particle correlations
- Particle spectra
- Collective flow

- Systematic dependencies of all the above with respect to
  - Collision energy
  - Centrality
  - Collision species

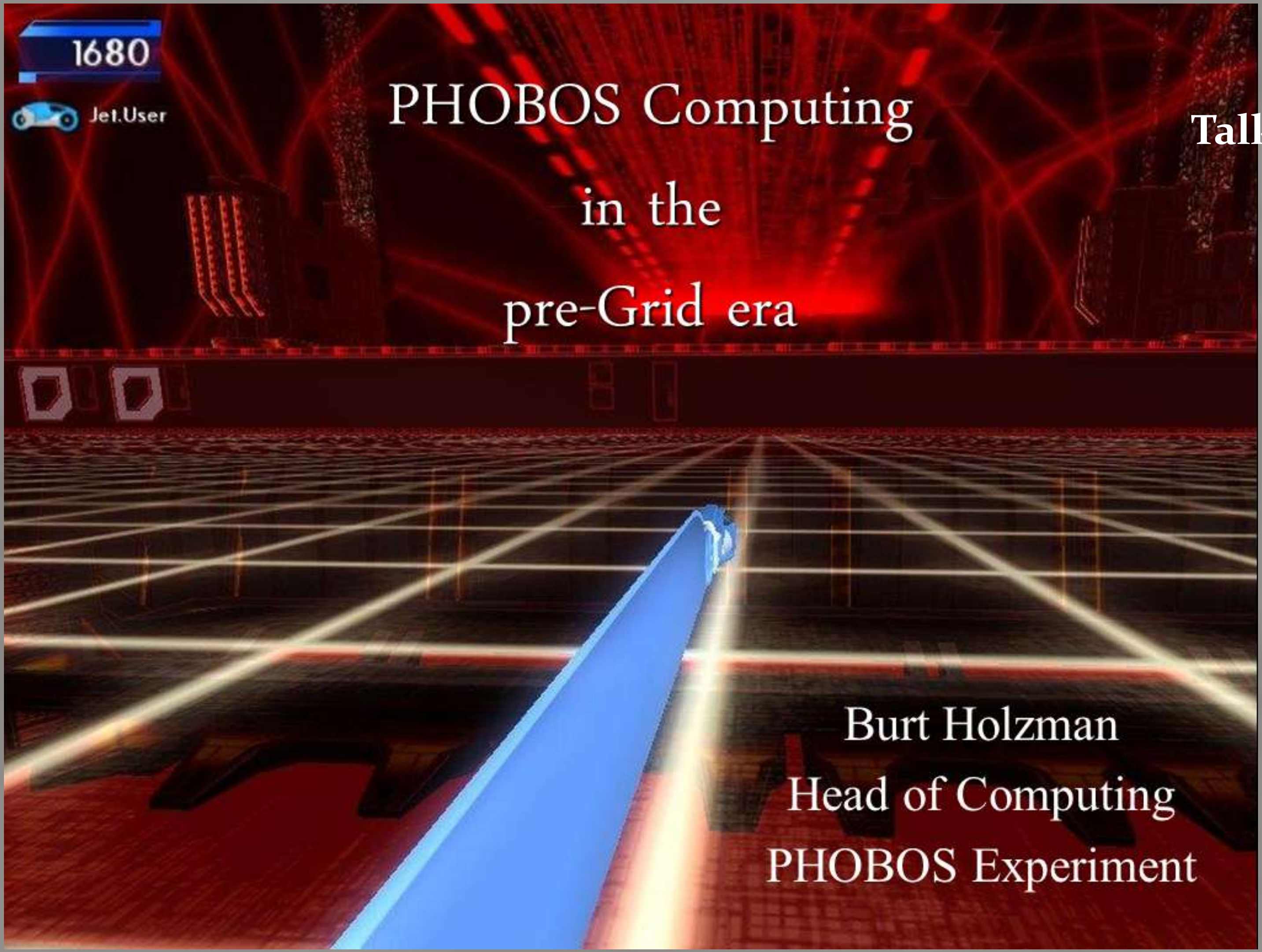
Total number of citable papers analyzed:	127
Total number of citations:	5054
Average citations per paper:	39.8
Renowned papers (500+)	1
Famous papers (250-499)	1
Very well-known papers (100-249)	10
Well-known papers (50-99)	11
Known papers (10-49)	37
Less known papers (1-9)	50
Unknown papers (0)	17
h-index	32

Citation data from INSPIRE

PHOBOS Computing

- 500 million events, ~100 kB each
- Aggregate data size: 50 TB/yr
- Data reconstructed on-the-fly at RHIC Computing Facility @ BNL

A look back at PHOBOS Computing



Talk from 2004

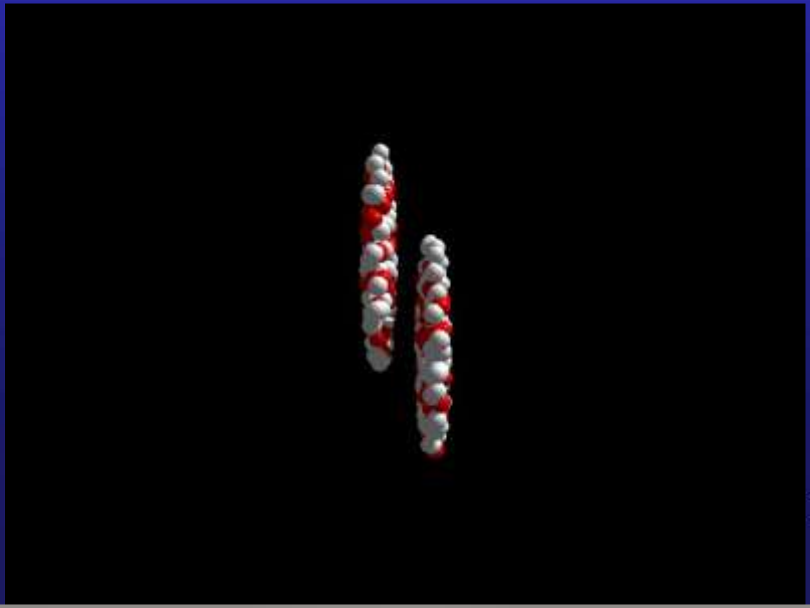
Collision Physics

We (HEP/Nuclear/RHIC) collide billions (and billions...) of

- electrons vs. positrons
- antiprotons vs. protons
- nuclei vs. nuclei
- ... and a whole lot more

Collision Physics

Collision physics is ideal for parallel computing: each collision ("event") is independent!



Animation courtesy of UrQMD group

Size of the PHOBOS Farm

pharm

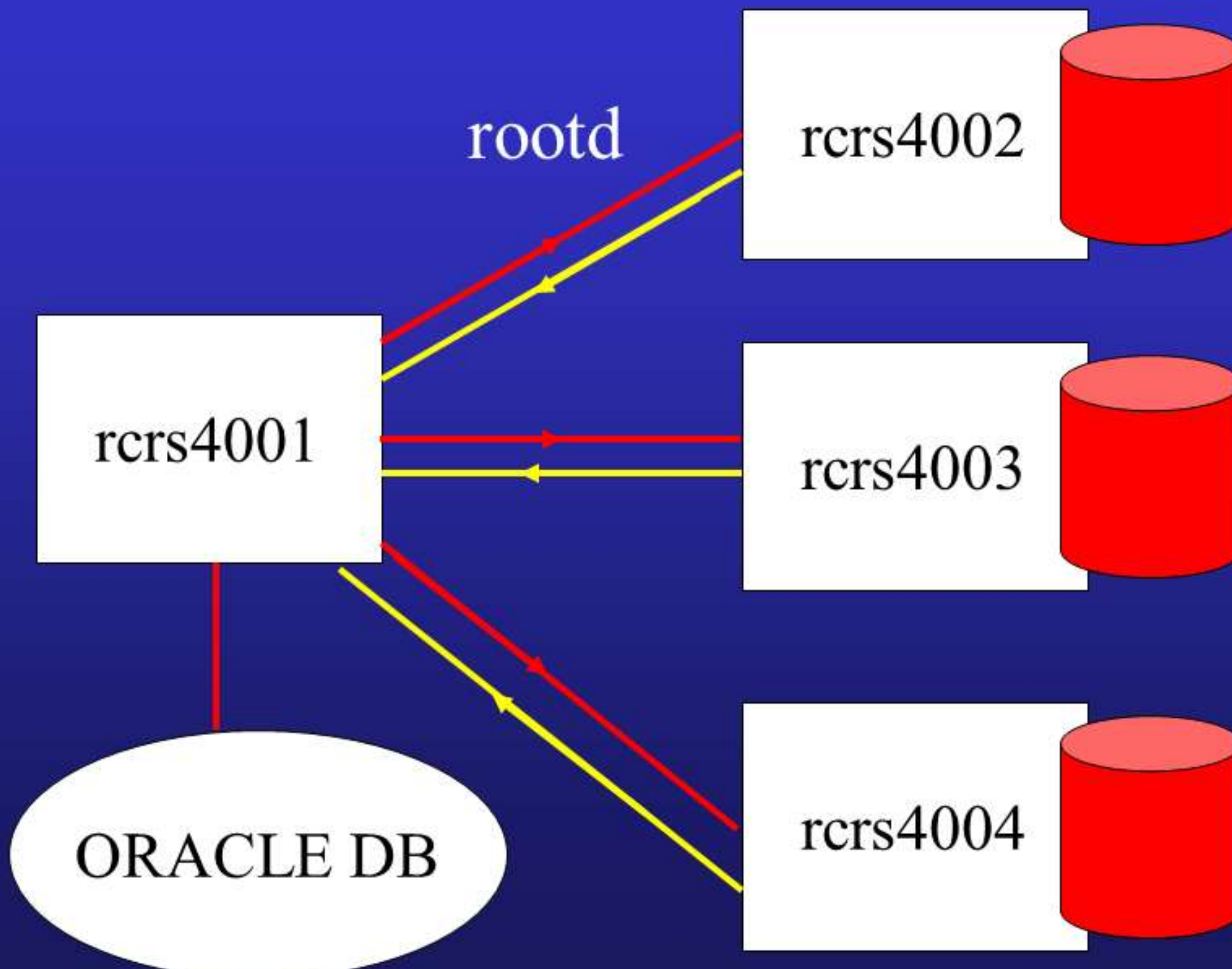
- 6 dual 733 MHz
- 6 dual 933 MHz
- 11 dual 1733 MHz

RHIC Computing Facility

- 30 dual 450 MHz
- 44 dual 800 MHz
- 63 dual 1000 MHz
- 26 dual 1400 MHz
- 98 dual 2400 MHz
- 80 dual 3060 MHz

Total: 728 cores

Distributed Disk



Distributed Disk: CatWeb

FileSets | Pools | Jobs | Adm

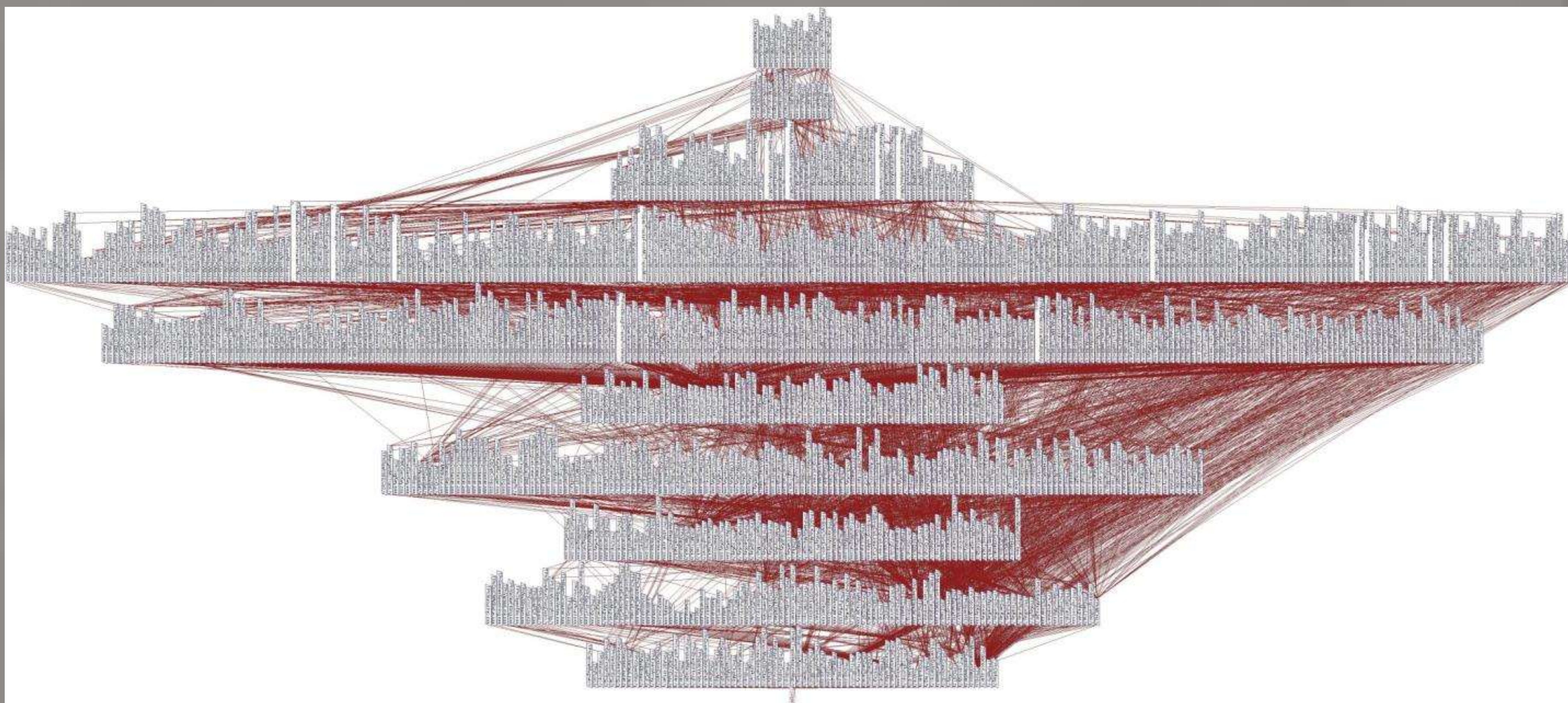
The Following Pools Exist:

88 Tb

Pool Name	Number of Disks	Total Disk Space (GB)	Ava e (GB)	Used (%)	Stat
rcf	946	81876	3027	60	Onli
pdev	8	739	524	29	Onli
pharm	63	5400	2061	62	Onli
pdev_2	8	734	732	0	Onli

PHOBOS Computing

PHOBOS Software Dependency graph

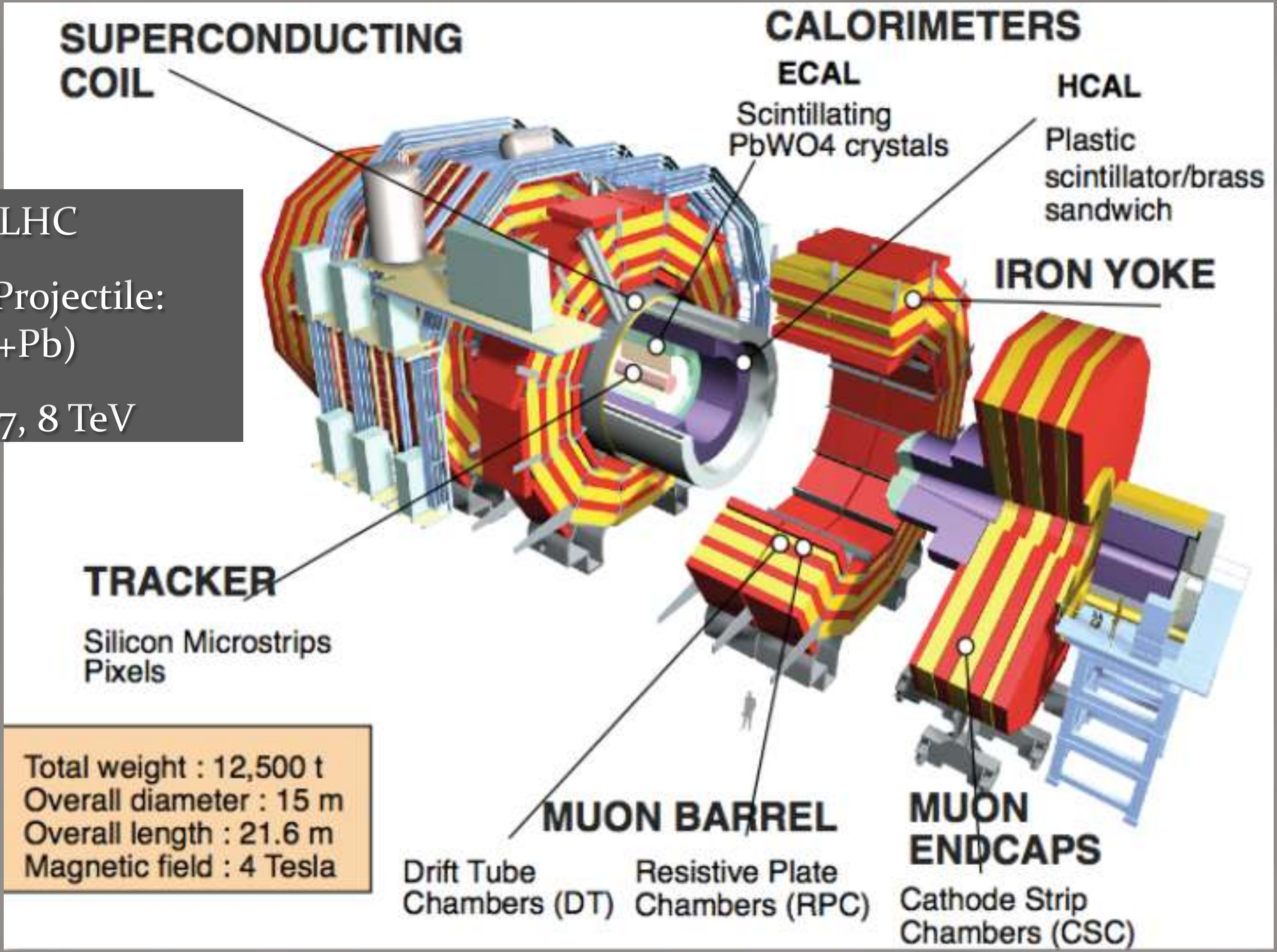


(or, what happens when a generation of physicists learns to write C++ for the first time…)

CMS Collaboration

CMS Experiment

- CMS @ LHC
- Target, Projectile: p, p (Pb+Pb)
- E_{BEAM} = 7, 8 TeV



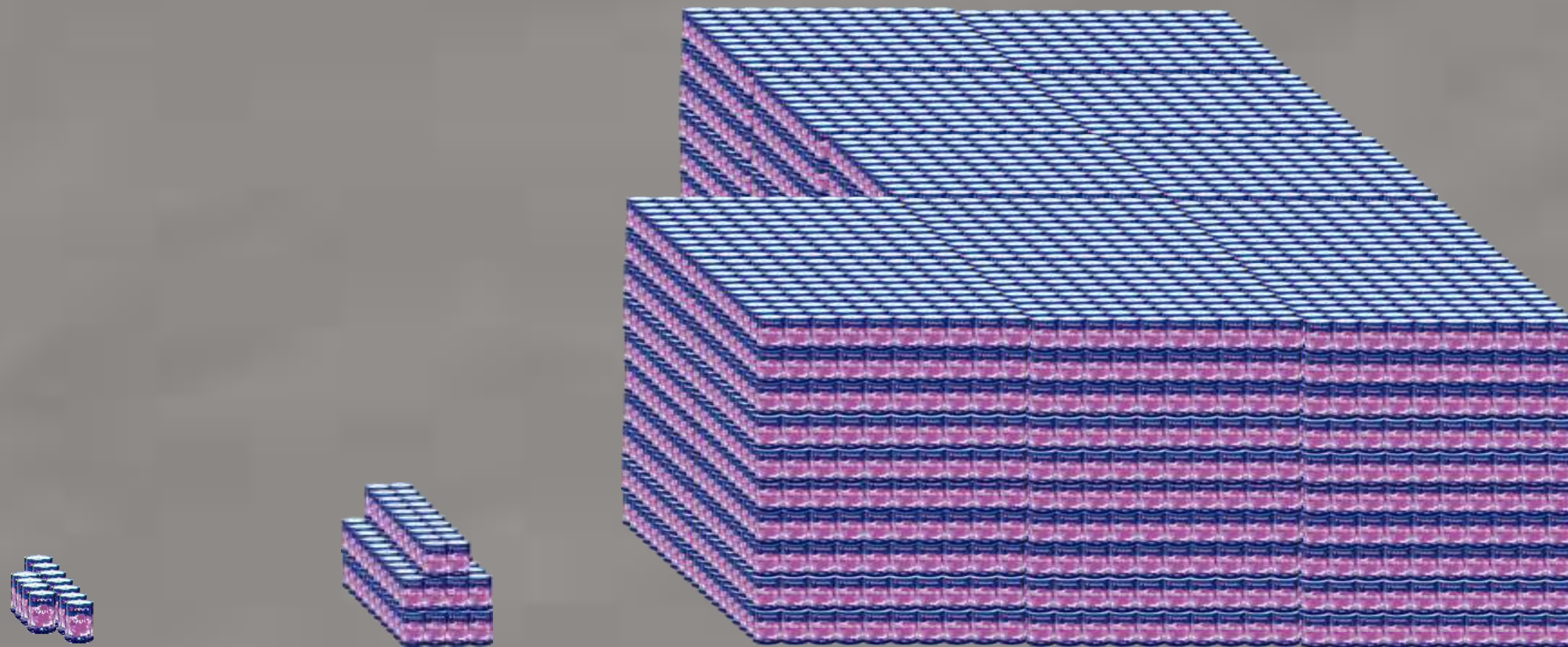
CMS Physics

- Could discover:
  - <u>**Higgs Boson**</u>
  - Supersymmetry
  - Extra Dimensions
  - Dark Matter
  - Dark Energy
  - and much, much more…

Total number of citable papers analyzed:	416
Total number of citations:	4647
Average citations per paper:	11.2
Renowned papers (500+)	1
Famous papers (250-499)	1
Very well-known papers (100-249)	8
Well-known papers (50-99)	5
Known papers (10-49)	74
Less known papers (1-9)	149
Unknown papers (0)	178
h-index	31

Citation data from INSPIRE

CMS Computing



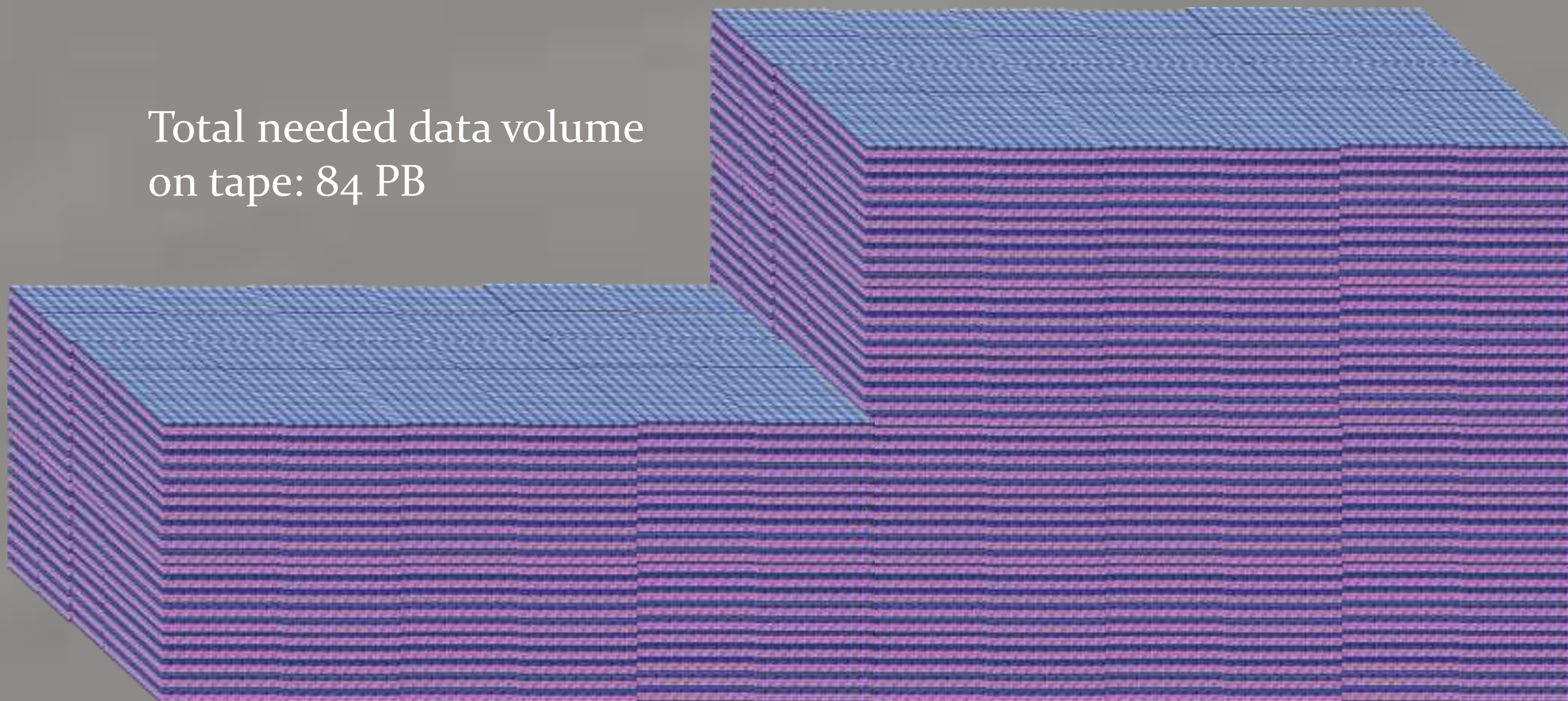
E917
5 TB/run

Phobos
50 TB/run

CMS
6 PB raw/run

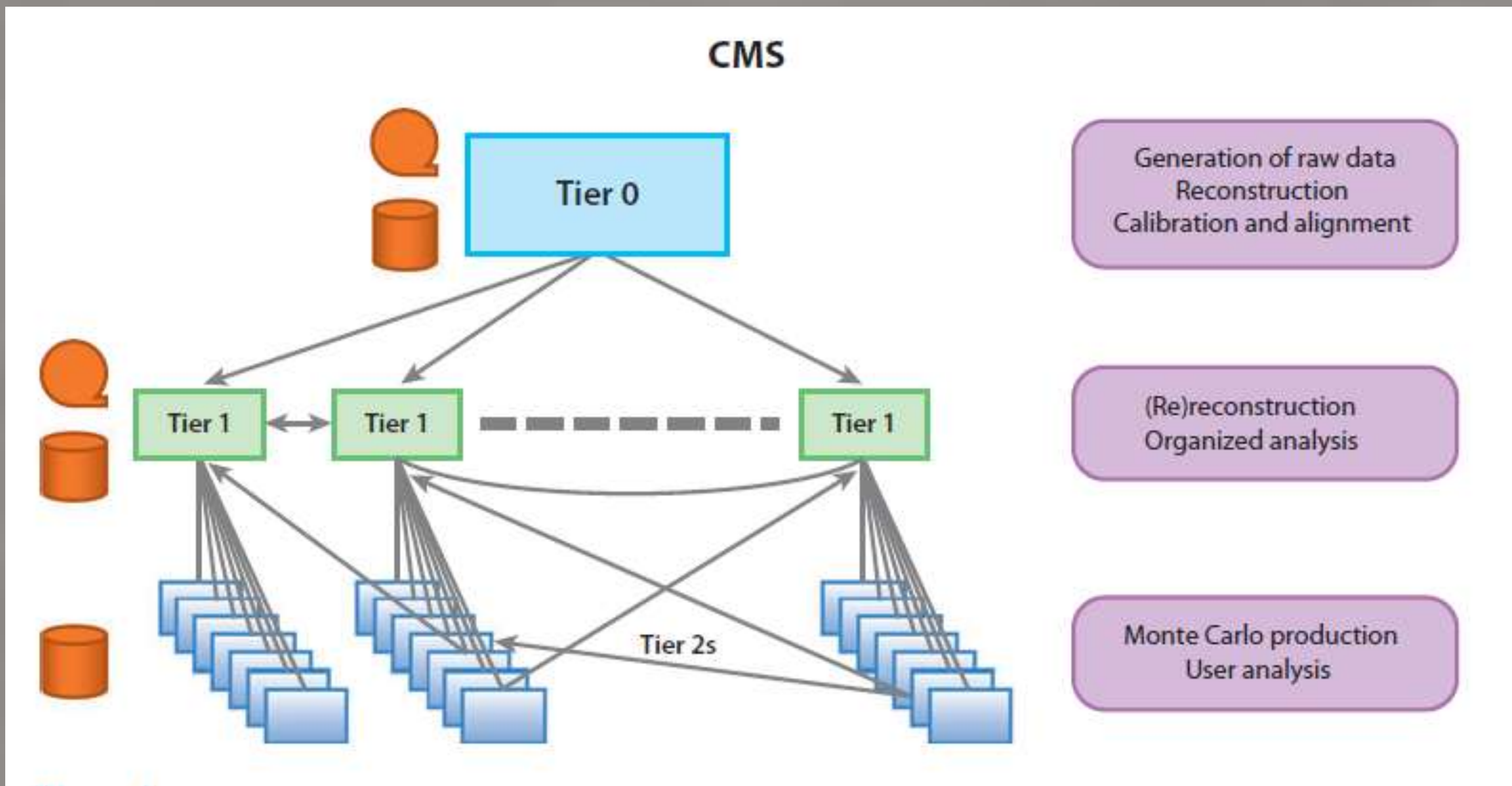
CMS Computing

Total needed data volume on tape: 84 PB

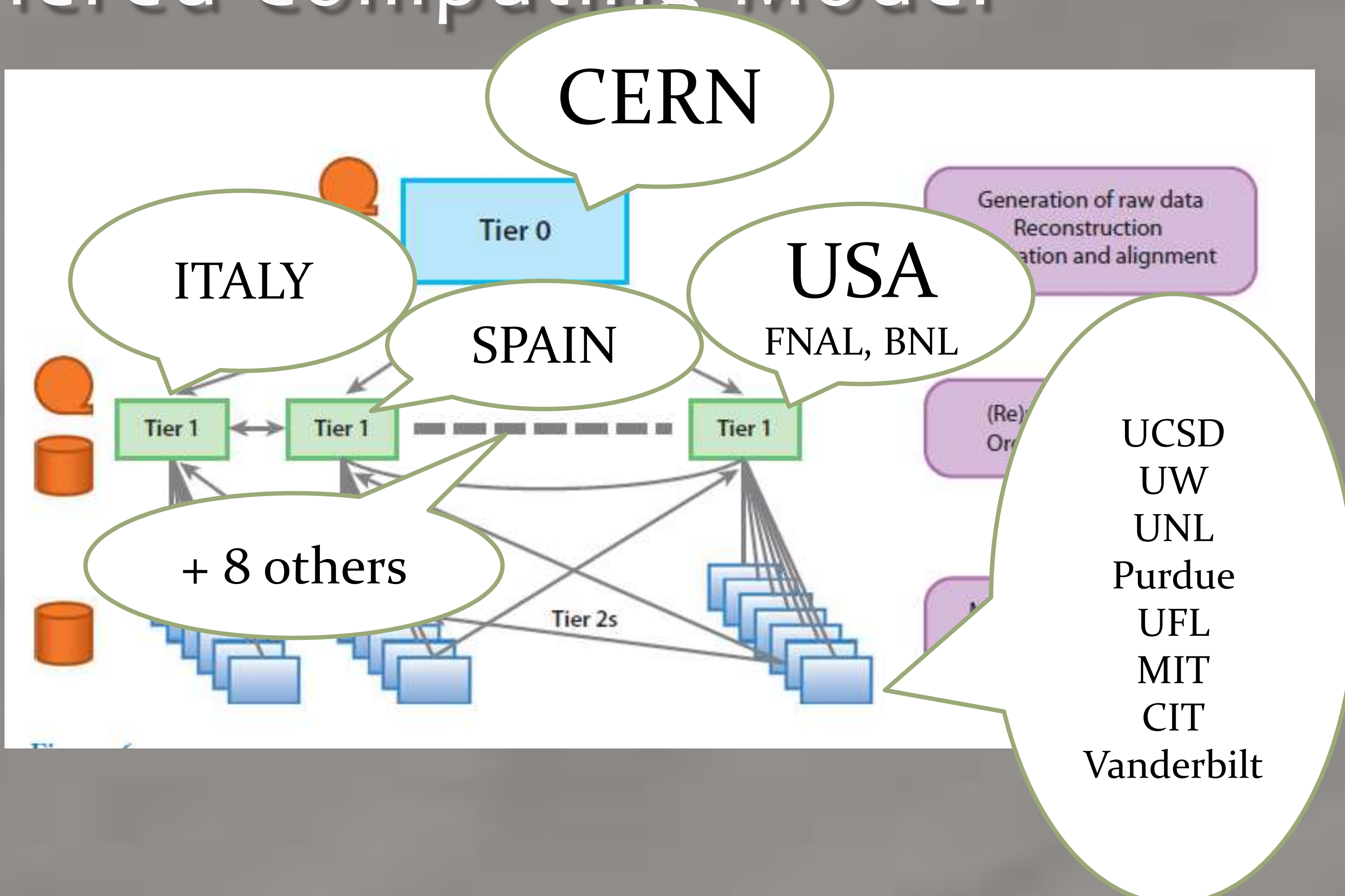


E917 Phobos

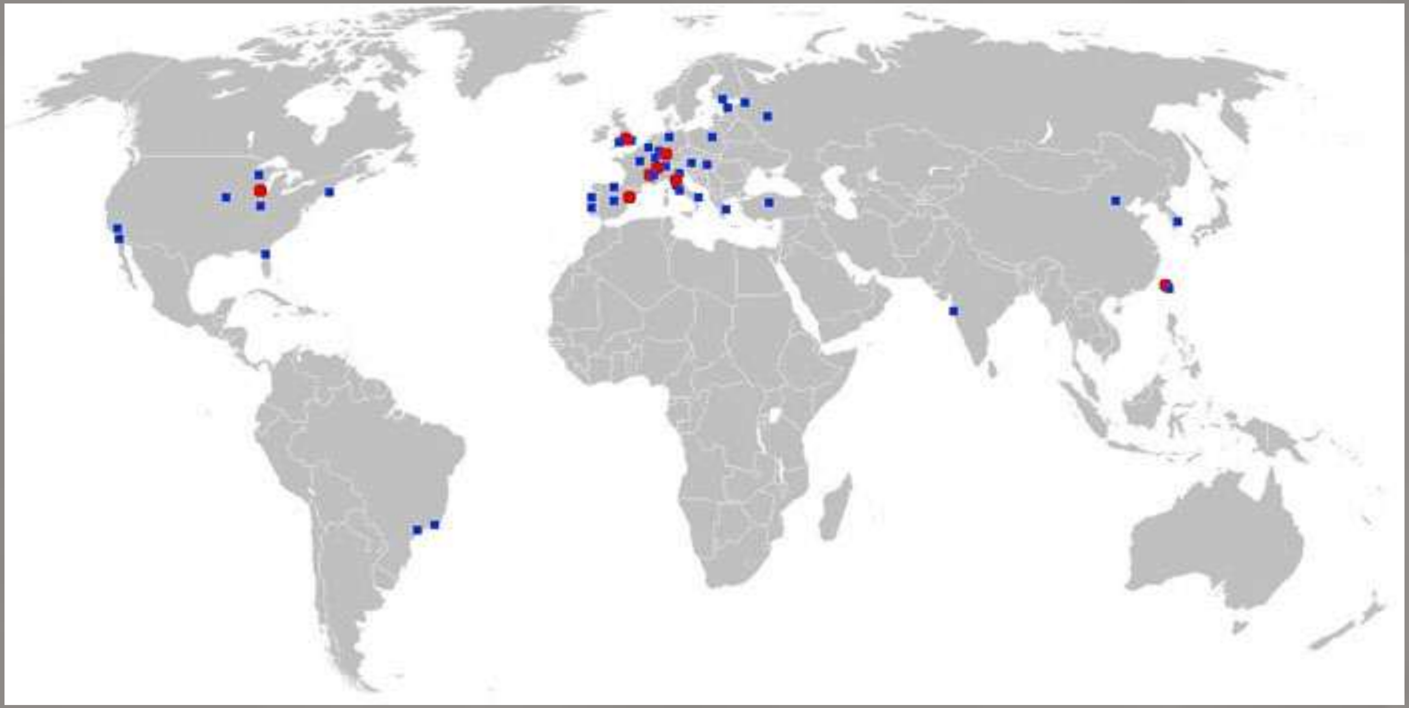
"MONARC" Tiered Computing Model



“MONARC”
Tiered Computing Model



CMS Computing Hierarchy

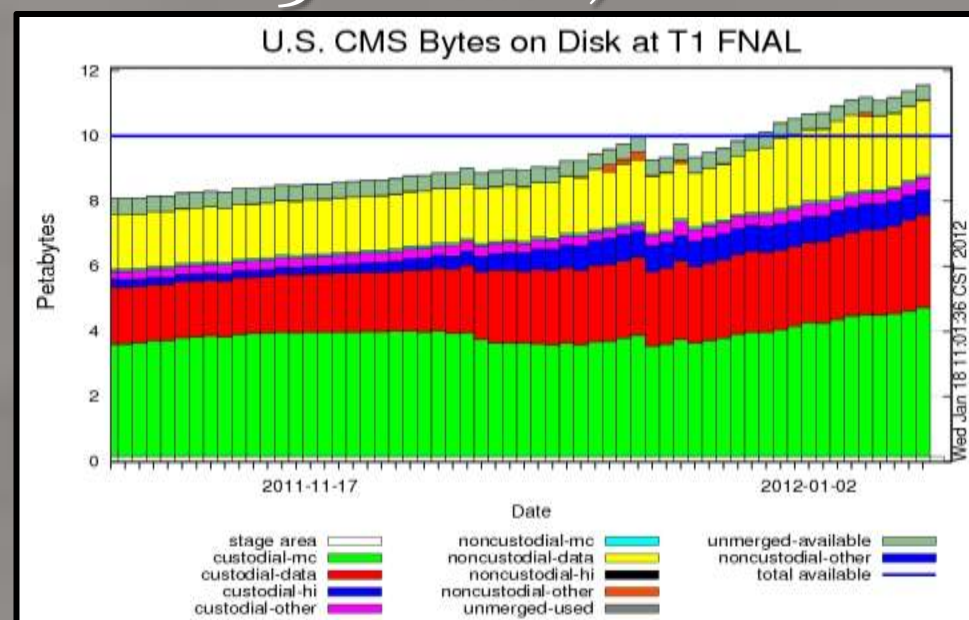


RED: Tier 1: Fermilab (US), ASGC (Taiwan), INFN (Italy), KIT (Germany), RAL (UK), PIC (Spain), IN2P3 (France)
BLUE: Tier 2: (roughly 50)

Fermilab is the only Tier 1 dedicated to CMS at 100%;
Consequently we are pledged to **40%** of the T1 computing share

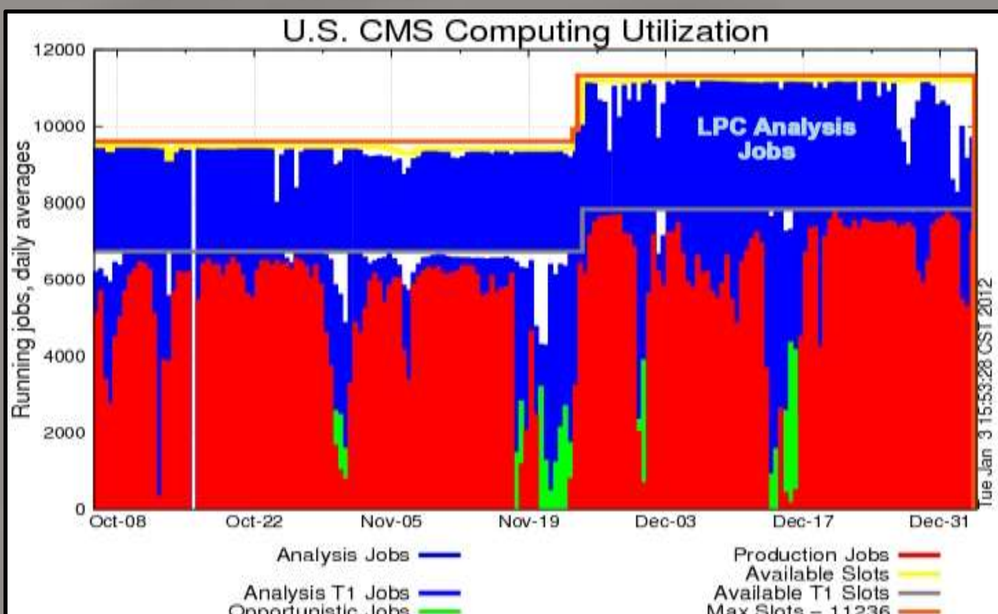
Fermilab Tier 1 for CMS

- Storage:
  - 15 Petabytes (PHOBOS: .088 PB)
  - Nexsan E60s (60 disks, 2-3 TB each)



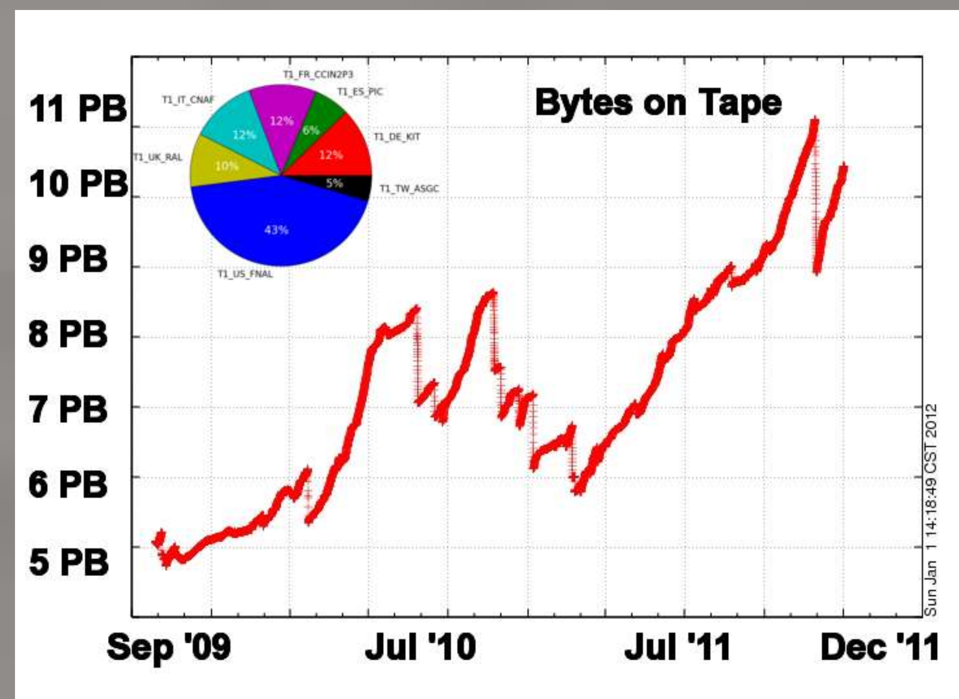
Fermilab Tier 1 for CMS

- Computing:
  - 3300 cores locally available
  - 8400 cores globally available



Fermilab Tier 1 for CMS

- Tape:
  - 21 Petabytes
  - STK SL8500 bots (10000 tapes each)



Fermilab Tier 1 Challenges

- Computing
  - Condor batch system: should scale to 100,000 jobs
- Disk
  - dCache: provides a flat namespace for distributed disk
  - Hadoop is proving a reliable alternative (if no off-line storage)
- Tape
  - Recent advances in media increase capacity to 5 TB/tape (1500 x bigger than in E917 days!)
- Network
  - "Data-intensive" HEP is still fairly CPU-intensive
  - 10 Gigabit networking is becoming commodity

How do we tie together all these distributed resources?

To the Grid!

"The Grid is an emerging infrastructure that will fundamentally change the way we think about – and use – computing. The word Grid is used by analogy with the electric power grid, which provides pervasive access to electricity"

[I. Foster & C. Kesselman, 1998]

We primarily use two Grids:
European Grid Initiative (EGI) and
Open Science Grid (OSG)

European Grid Initiative

The **European Grid Infrastructure (EGI)** delivers integrated computing services to European researchers, driving innovation and enabling new solutions to answer the big questions of tomorrow.

EGI's mission is to allow researchers of all fields to make the most out of the latest computing technologies for the benefit of their research.

EGI is a federation of over 350 resource centres and coordinated by EGI.eu, a not-for-profit foundation created to manage the infrastructure on behalf of its participants: National Grid Initiatives (NGIs) and European Intergovernmental Research Organisations (EIROs). EGI.eu is governed by a Council of 35 participant countries and institutions.

Open Science Grid

The Open Science Grid (OSG) advances science through open distributed computing. The OSG is a multi-disciplinary partnership to federate local, regional, community and national cyberinfrastructures to meet the needs of research and academic communities at all scales.

The **O**pen **S**cience **G**rid (**OSG**) provides provide common service and support for resource providers and scientific institutions using a distributed fabric of high throughput computational services. The OSG does not own resources but provides software and services to users and resource providers alike to enable the opportunistic usage and sharing of resources. The OSG is jointly funded by the Department of Energy and the National Science Foundation.

Open Science Grid

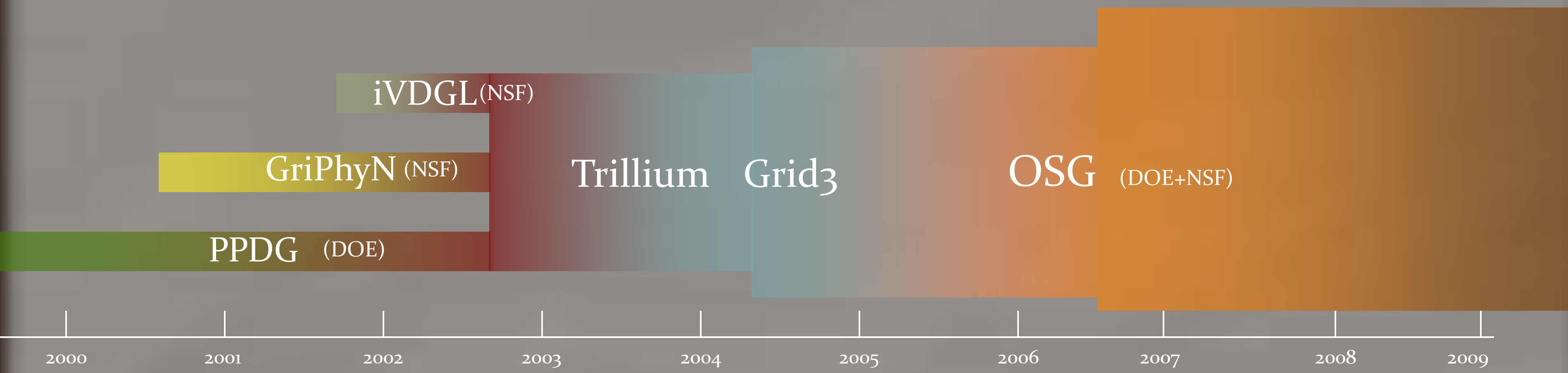
"Transform processing and data intensive science through a cross-domain self-managed national distributed cyber-infrastructure that brings together campus and community infrastructure and facilitating the needs of Virtual Organizations (VO) at all scales."

[M. Livny, 2007]

VO is nearly analogous to experiment – CMS is an OSG VO

Fermilab is a key stakeholder on OSG, so I'm focusing on OSG in this talk

Birth of the OSG



OSG “born” about 2005
(when I joined CMS)

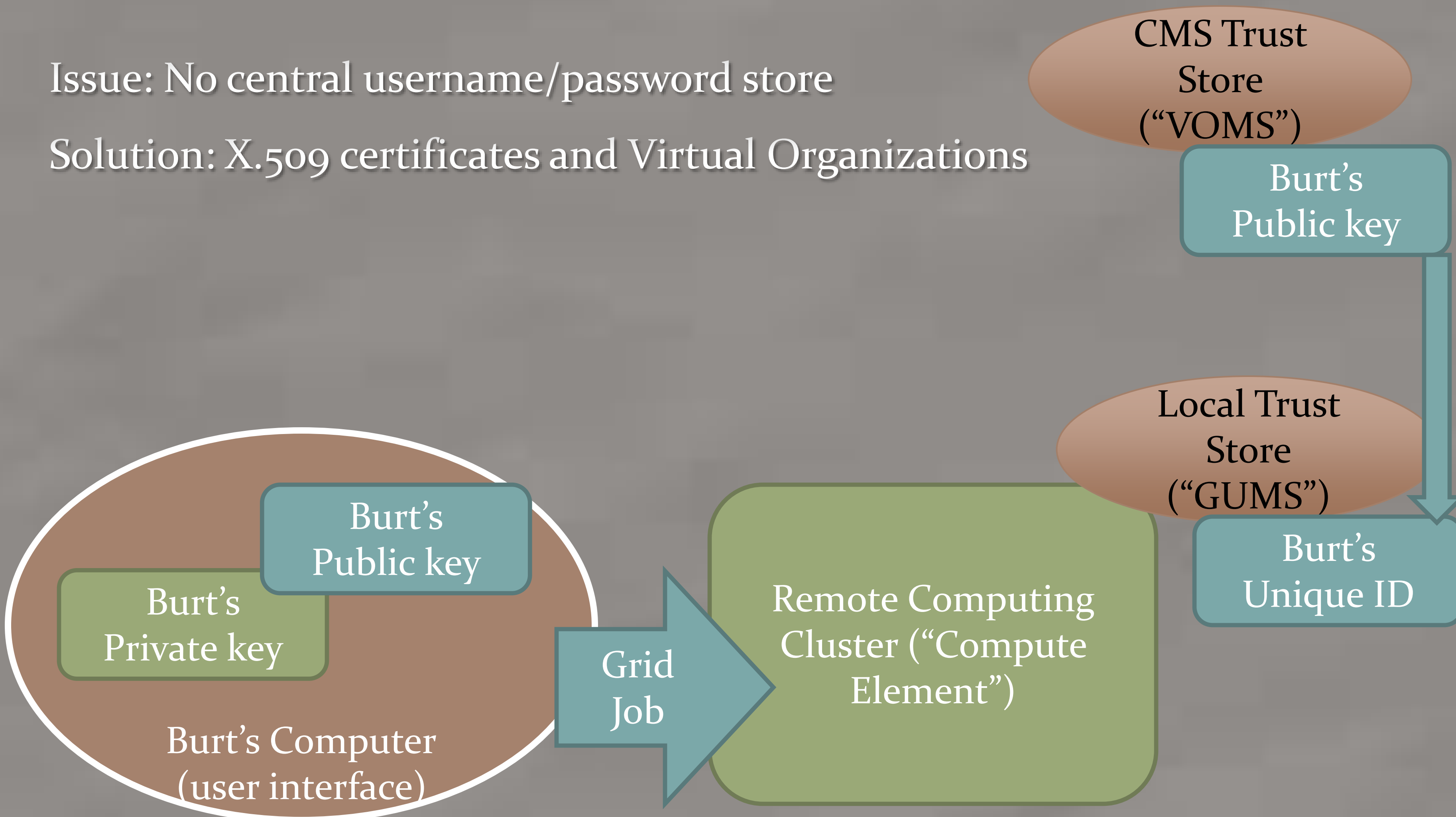
What is OSG, really?

- More than 100 heterogeneous clusters of Linux machines
- Handful of different storage solutions with a common protocol
- Common interface to computing and storage
- Opportunistic use
- Interoperability with other grids

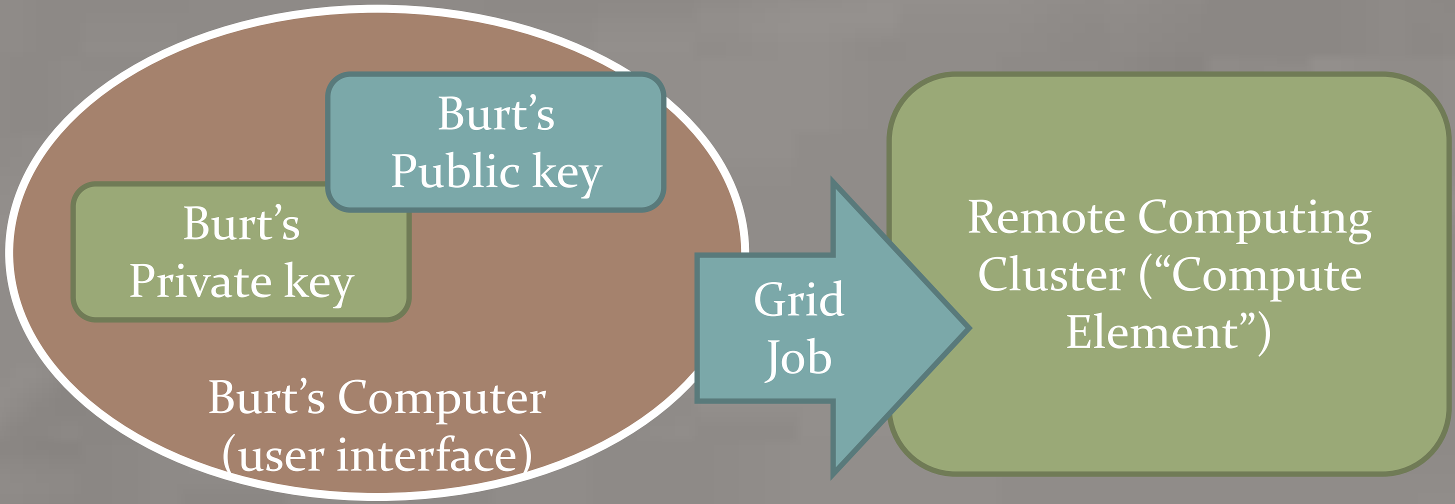
Challenges of Grid Computing: Authorization / Authentication

Issue: No central username/password store

Solution: X.509 certificates and Virtual Organizations



Challenges of Grid Computing: Common Interfaces



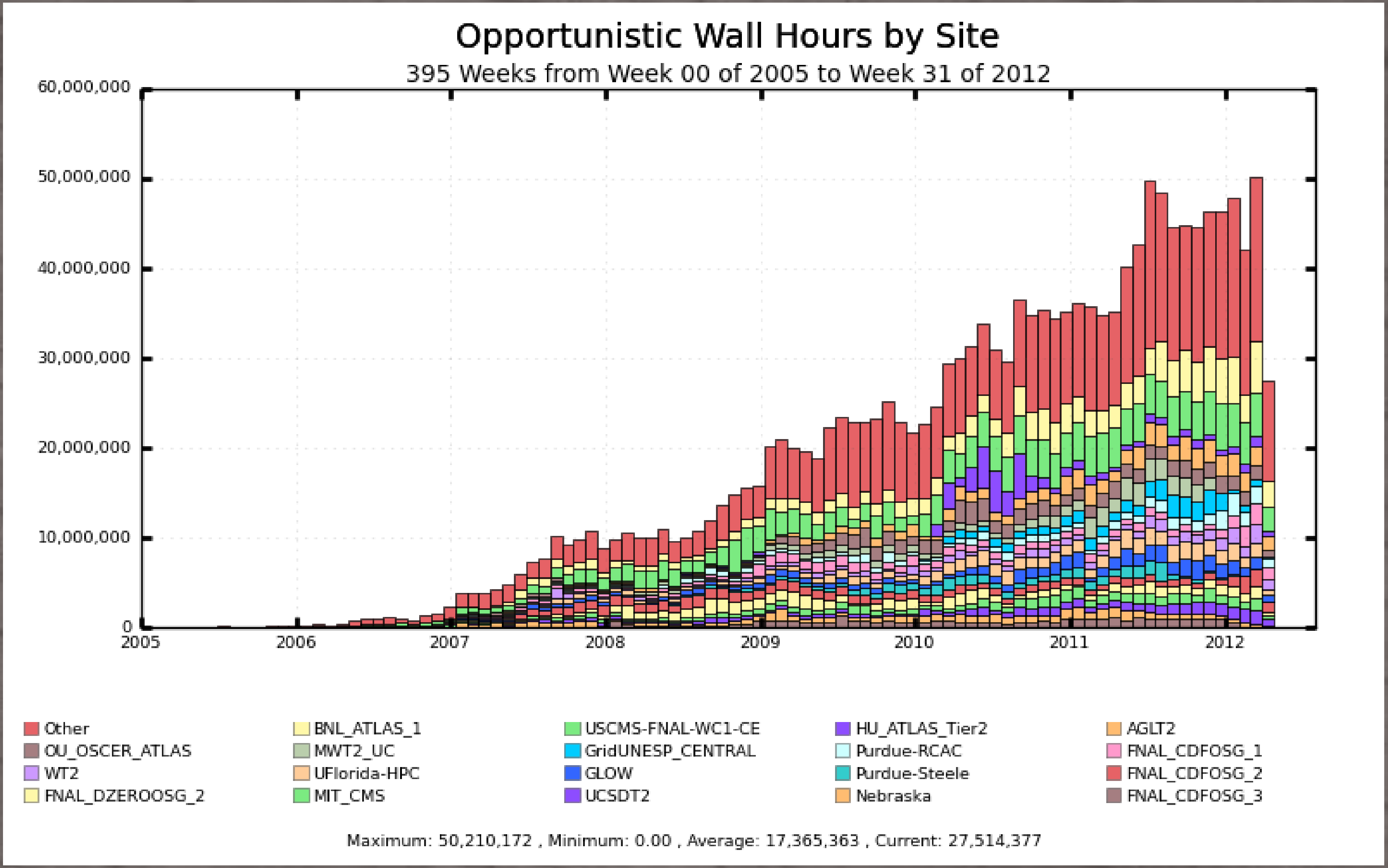
Issue: Compute Element has to communicate with common local schedulers (Condor, PBS, SGE, LSF …)

Solution: Globus Toolkit (Globus Consortium / ANL)

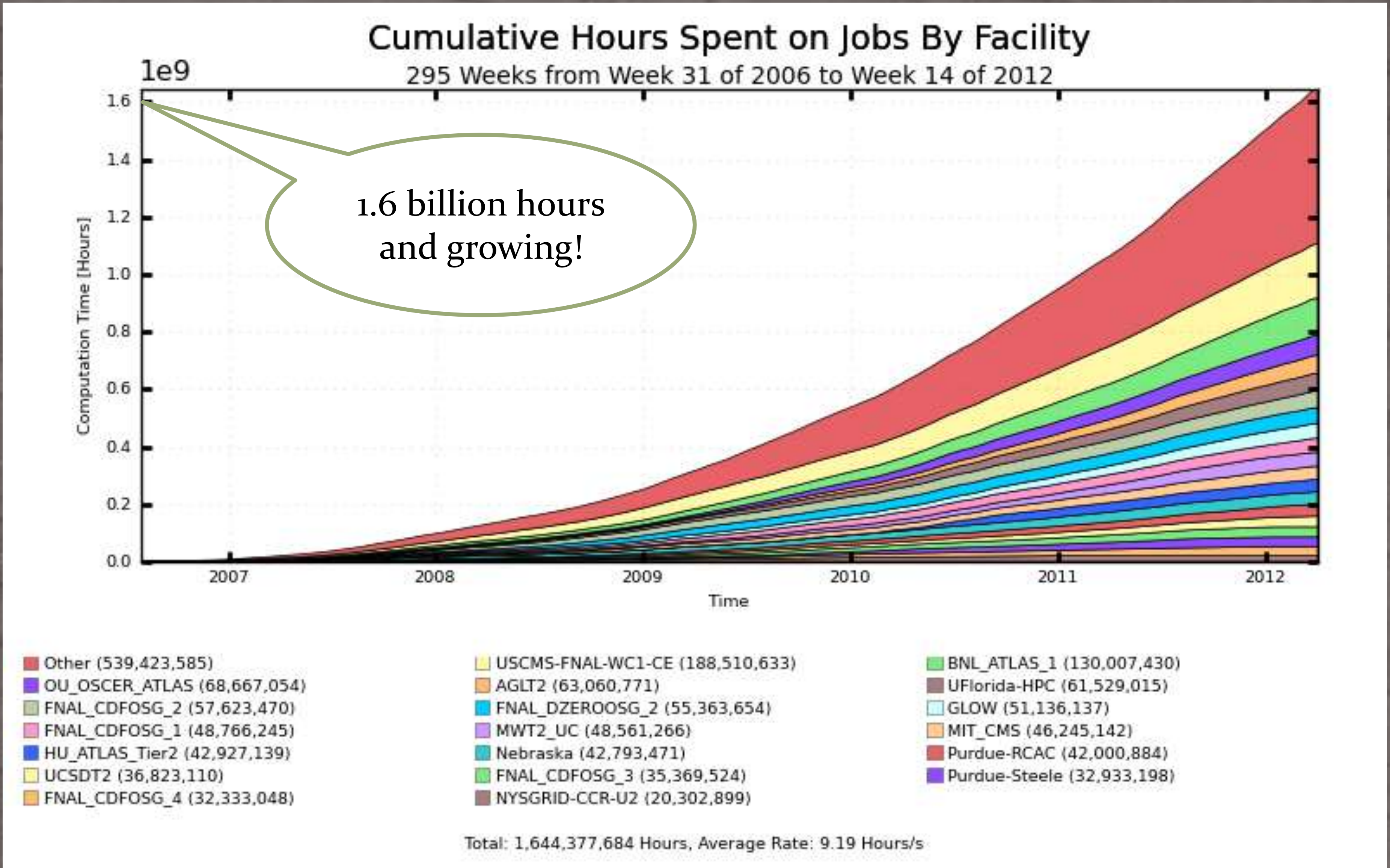
Issue: User interface needs to communicate to various Compute Elements

Solution: Condor-G (Condor Team @ UW-Madison)

Does it work?



Opportunistic Wall Hours by Site
395 Weeks from Week 00 of 2005 to Week 31 of 2012
60,000,000
50,000,000
40,000,000
30,000,000
20,000,000
10,000,000
0
2005
2006
2007
2008
2009
2010
2011
2012
Other
BNL_ATLAS_1
USCMS-FNAL-WC1-CE
HU_ATLAS_Tier2
AGLT2
OU_OSCER_ATLAS
MWT2_UC
GridUNESP_CENTRAL
Purdue-RCAC
FNAL_CDFOSG_1
WT2
UFlorida-HPC
GLOW
Purdue-Steele
FNAL_CDFOSG_2
FNAL_DZEROOSG_2
MIT_CMS
UCSDT2
Nebraska
FNAL_CDFOSG_3
Maximum: 50,210,172 , Minimum: 0.00 , Average: 17,365,363 , Current: 27,514,377

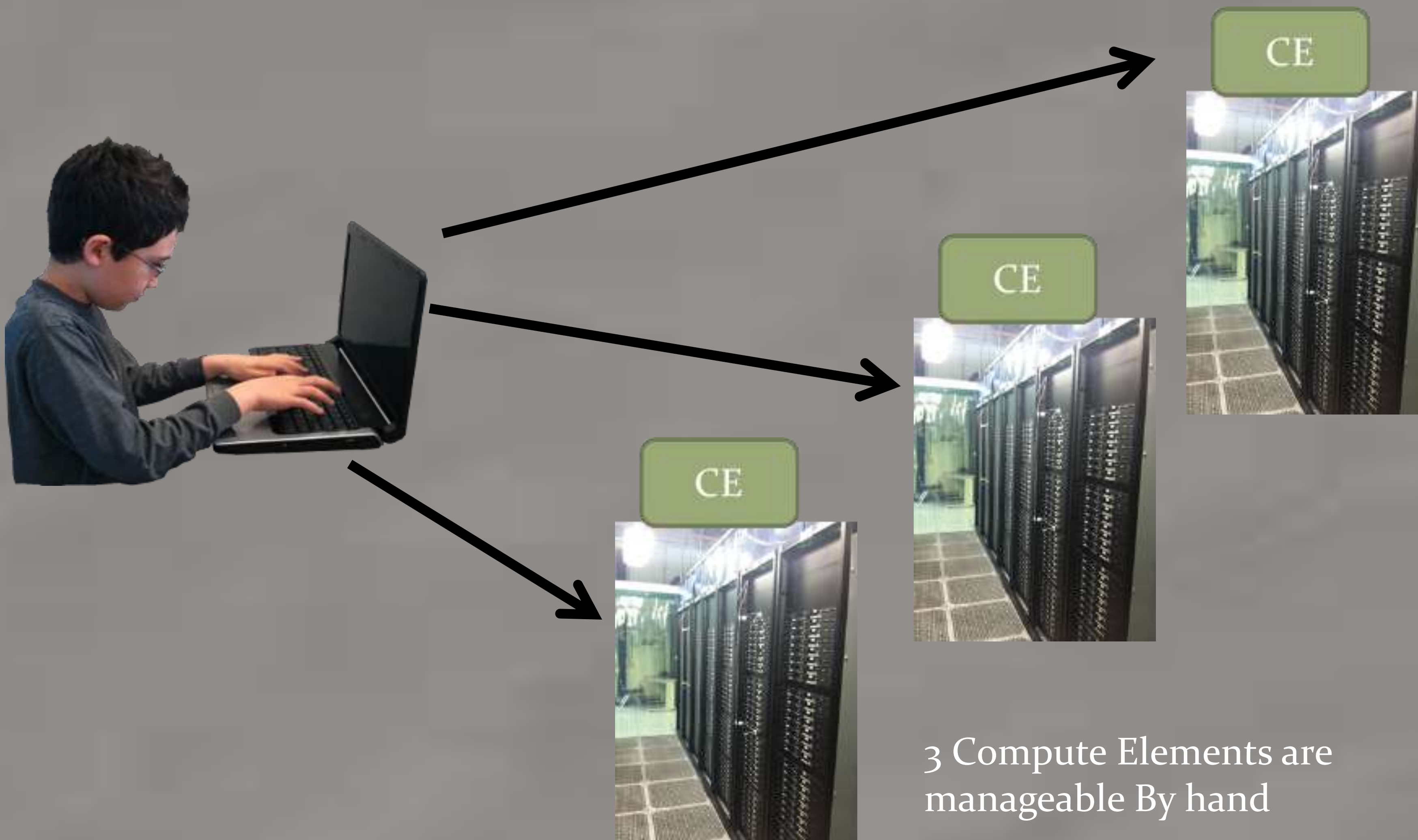


Cumulative Hours Spent on Jobs By Facility
295 Weeks from Week 31 of 2006 to Week 14 of 2012
1e9
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Computation Time [Hours]
2007
2008
2009
2010
2011
2012
Time
1.6 billion hours
and growing!
Other (539,423,585)
USCMS-FNAL-WC1-CE (188,510,633)
BNL_ATLAS_1 (130,007,430)
OU_OSCER_ATLAS (68,667,054)
AGLT2 (63,060,771)
UFlorida-HPC (61,529,015)
FNAL_CDFOSG_2 (57,623,470)
FNAL_DZEROOSG_2 (55,363,654)
GLOW (51,136,137)
FNAL_CDFOSG_1 (48,766,245)
MWT2_UC (48,561,266)
MIT_CMS (46,245,142)
HU_ATLAS_Tier2 (42,927,139)
Nebraska (42,793,471)
Purdue-RCAC (42,000,884)
UCSDT2 (36,823,110)
FNAL_CDFOSG_3 (35,369,524)
Purdue-Steele (32,933,198)
FNAL_CDFOSG_4 (32,333,048)
NYSGRID-CCR-U2 (20,302,899)
Total: 1,644,377,684 Hours, Average Rate: 9.19 Hours/s

Does it work?

Yes, but it could work better...

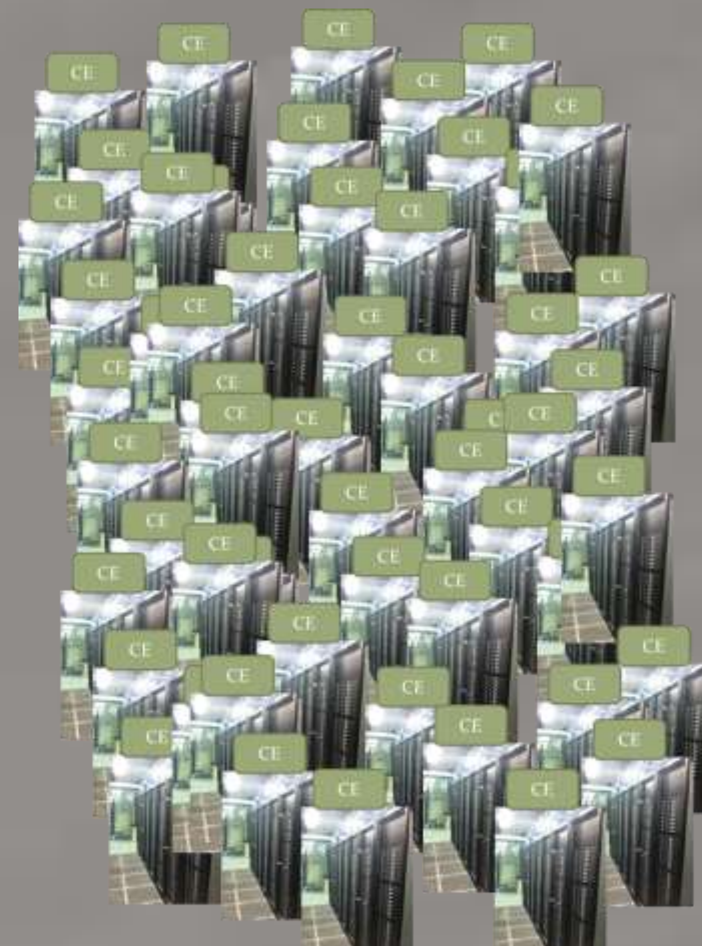
Challenges of Grid Computing: Distributed Compute Resources



3 Compute Elements are manageable By hand

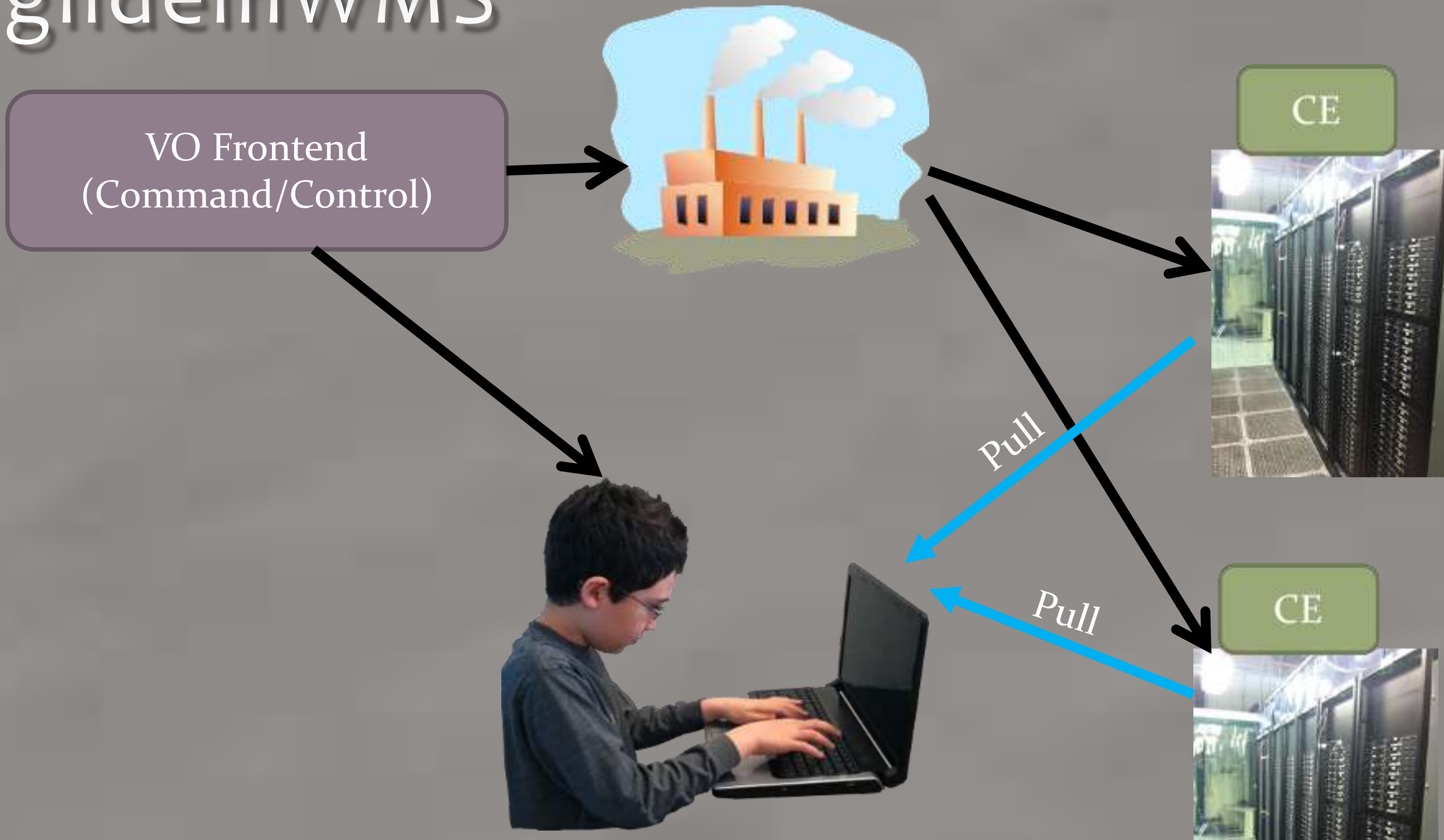
Challenges of Grid Computing: Distributed Compute Resources

?



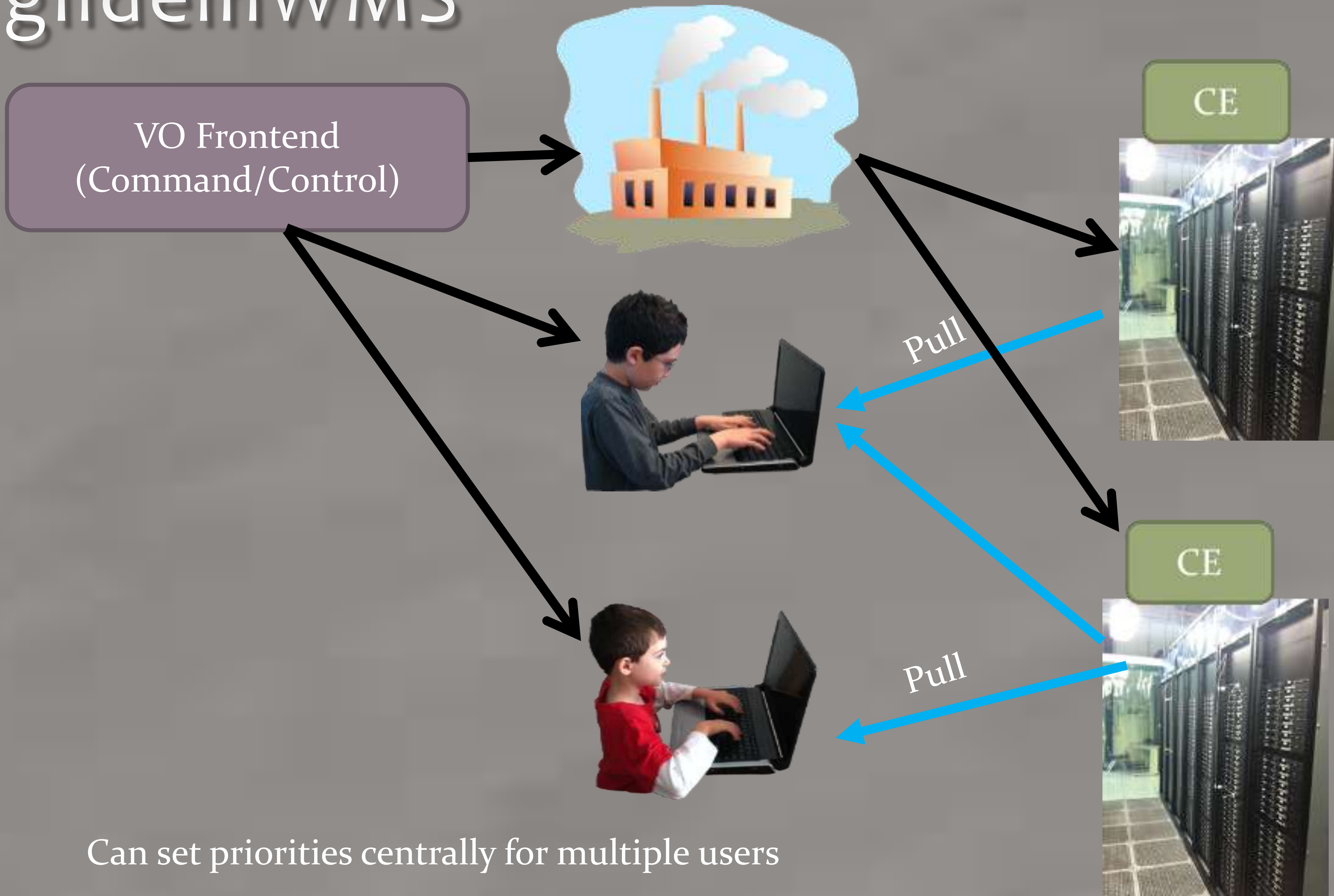
We need middleware – specifically
a Workload Management System
(and more specifically, “glideinWMS”)

glideinWMS



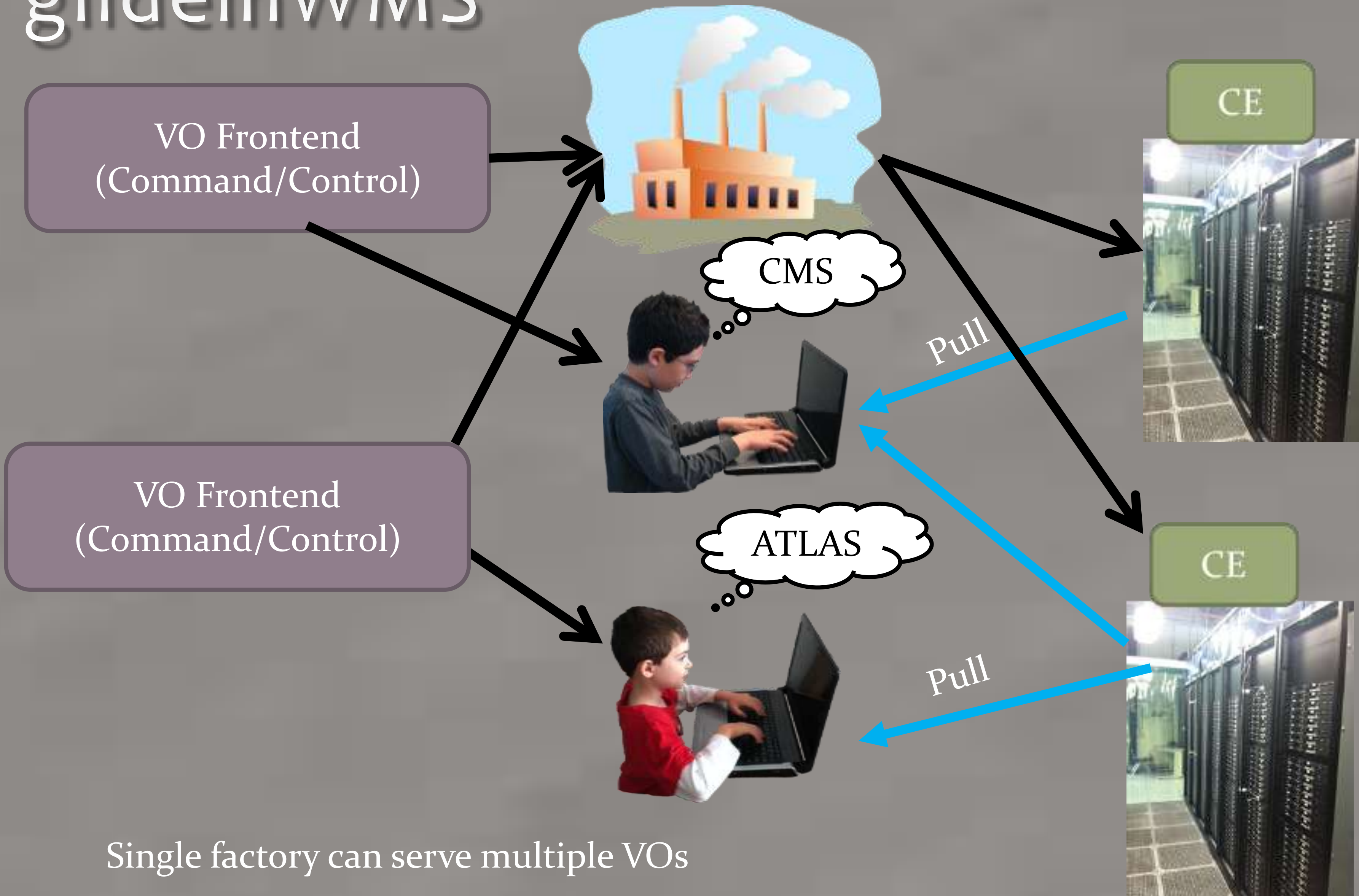
All "grid" knowledge remains at the factory
Grid looks like a Condor batch system to the user
Jobs run inside a container which protect user against bad nodes

glideinWMS



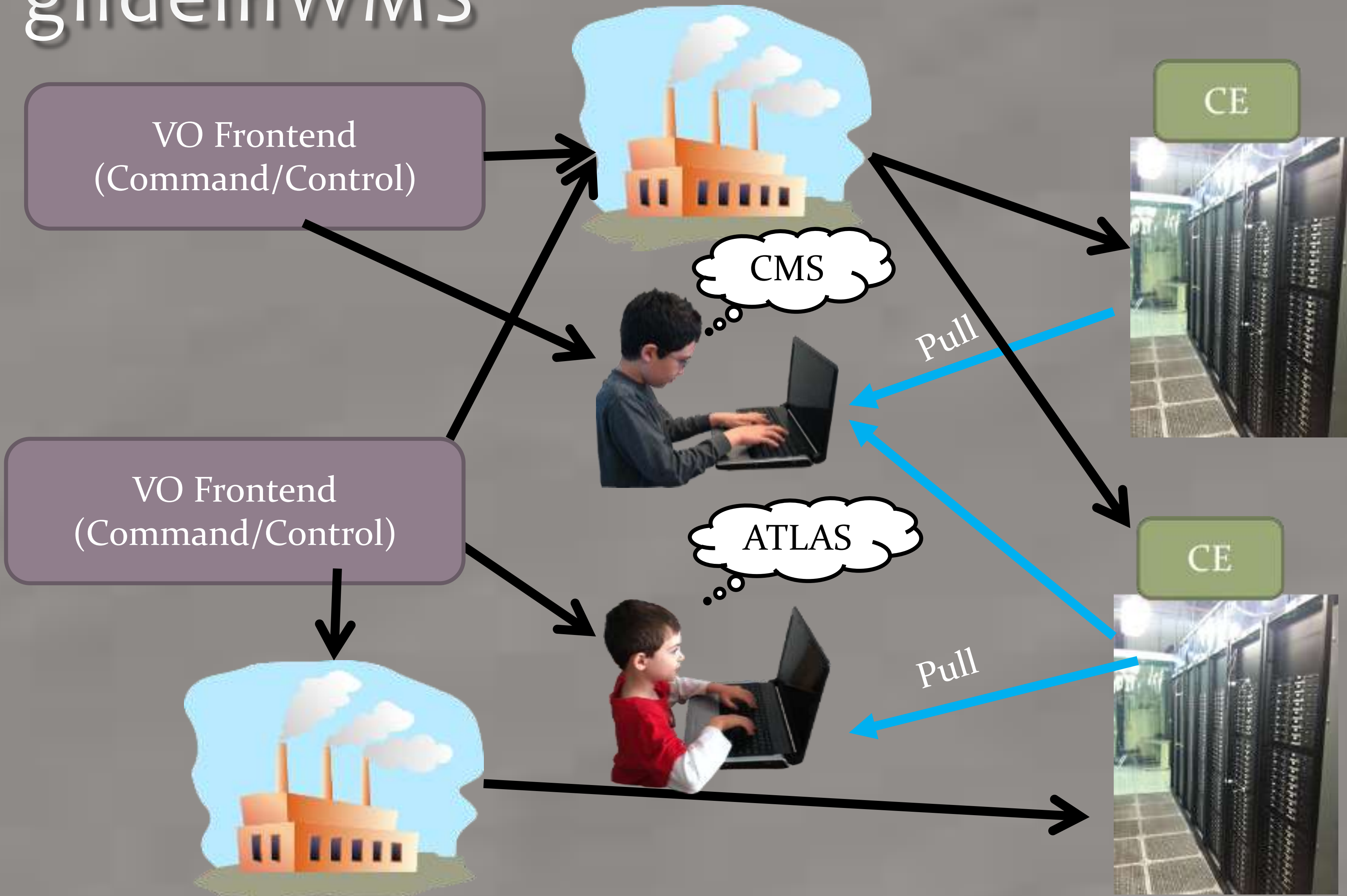
Can set priorities centrally for multiple users

glideinWMS



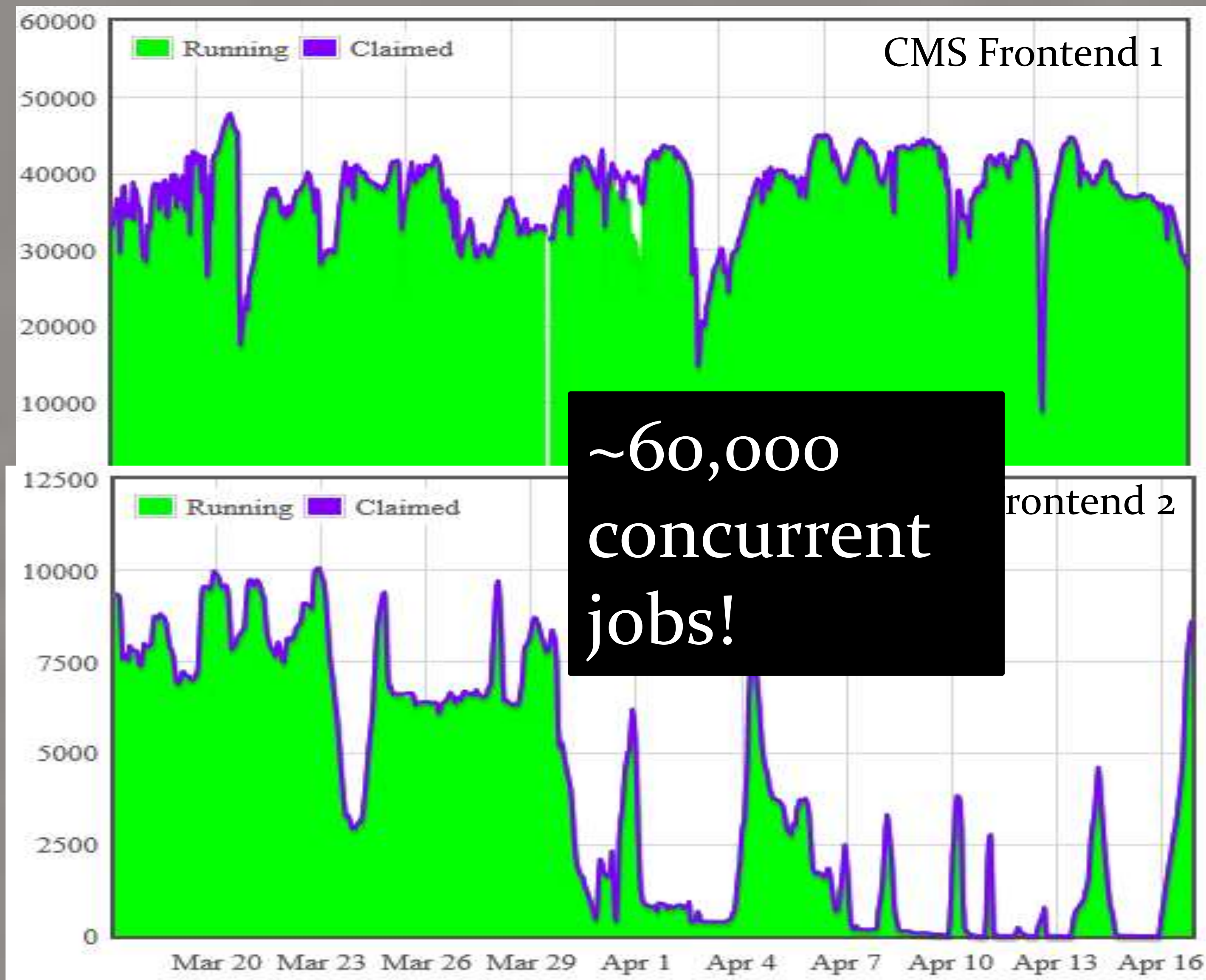
Single factory can serve multiple VOs

glideinWMS

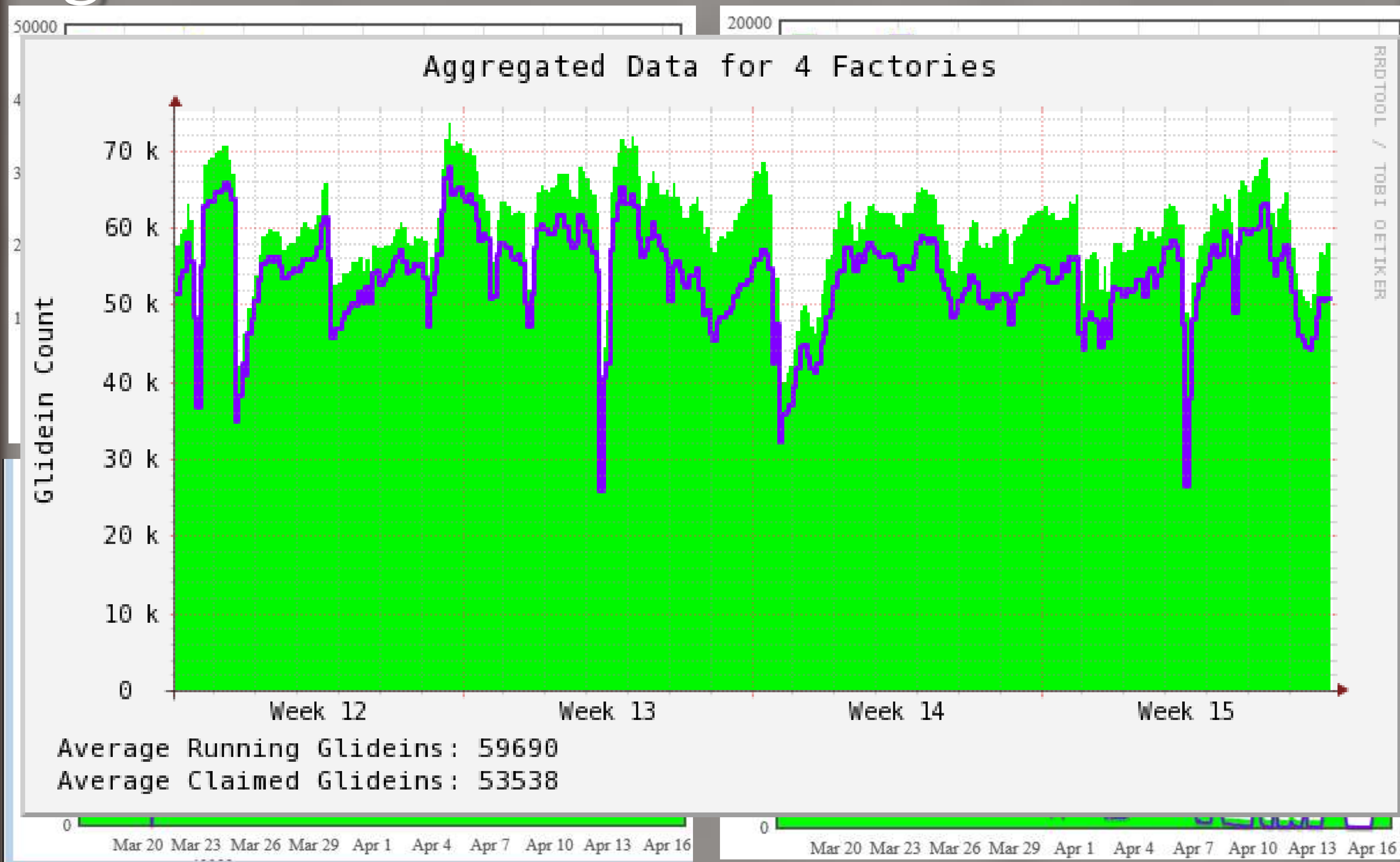


VO Frontend can talk to multiple factories

glideinWMS – CMS worldwide



glideinWMS results



Quo vadis? My wild guesses

- Data-intensive science gets more data-intensive – a few examples

SKA will acquire 1000 PB/day
(but "only" store few PB/day)

nature genetics

Genomics goal: full human DNA sequencing for $1000 (~ 1 GB/person)

Office of Science and Technology Policy
Executive Office of the President
New Executive Office Building
Washington, DC 20502

FOR IMMEDIATE RELEASE
March 29, 2012

Contact: Rick Weiss 202 456-6037 rweiss@ostp.eop.gov
Lisa-Joy Zgorski 703 292-8311 lisajoy@nsf.gov

OBAMA ADMINISTRATION UNVEILS "BIG DATA" INITIATIVE:
ANNOUNCES $200 MILLION IN NEW R&D INVESTMENTS

Even the government is paying attention!

Quo vadis? My wild guesses

- Computing
  - Nominal processor speed increase
  - Core density continues to increase (64 available now – 1024 coming before you know it)
- Storage
  - Solid State Drive prices approach traditional Hard Drive
  - Price per TB continues to fall after recovery from Thai flooding
- New class of challenges discovered when we scale up by another order of magnitude

Thank you

- DOE, NSF, and European funding agencies
- UIC Community
- Collaborators of past, present, and future
- My “volunteers” to model as users

Kyle Alex